

A logical Bayesian framework to dynamically compose a modular inductive model of a proteomics experiment

Authors: [Kurt De Grave, Jan Ramon]

Intro: (max 120 words)

Machine learning has been successfully applied to proteomics [Kelchtermans 2014] to model isolated subprocesses, intermediaries, or outcomes of proteomics experiments, such as enzymatic cleavage or intensity prediction. While successful in their own right, these models are brittle. Their use requires caution: you better make sure – by hand – that the model was trained on data of experiments that match yours.

Fortunately, HUPO-PSI controlled vocabularies and structured formats ensure consistent semantics between conforming documents. This formality enables inference over data that would otherwise not be possible or safe. Logic inference in turn can solve queries about the compatibility of pretrained models with the experiment at hand.

Methods (max 120 words)

We aim to obtain a comprehensive picture of the experimental process with a composite probabilistic model. We adopt the framework of logical Bayesian networks (LBN) [Fierens et al., 2005], which combine the properties of logical programs and Bayesian networks. In particular, a logic program is used to specify the random variables of interest and the Bayesian network defining their joined probability distribution. We construct a PSI-based LBN model describing the several domains relevant for proteomics, taking a modular approach as mandated by the need to accommodate different labs and experimental protocols. Labs can reconfigure the network to fit their experimental protocols. A GUI prototype is available to draw the experimental mass flow network.

Novel aspect (max 20 words)

A probabilistic logical model for proteomics.